DOCUMENT RESUME

ED 164 622 TM 008 337

AUTHOR Holmes, Mark
TITLE Report on the Intermediate Evaluation Project.
PUB DATE 15 Sep 77
NOTE 23p.; For related document, see TM 008 085

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Academic Achievement; *Educational Assessment; English; Foreign Countries; French; *Grade 8; Junior High Schools; Performance Factors; Problem Solving; Secondary School Mathematics; Secondary School Science; *State Surveys; *Student Attitudes; *Testing Programs; Test Results; Writing
IDENTIFIERS Canada; Ontario

ABSTRACT

Statistical data and comments describe the results of the 1977 eighth grade testing program in Ontario, Canada. The tests here: Mathematics (in French and English), Problem Solving (in French and English); Science (in English only), Attitudes, French comprehension, and Written English. Table 1 presents the numbers of students, by school district, taking each test. Tables 2 through 7 show the percent of students attaining the three achievement levels of competence, honors, and high honors, for each examination. Table 8 gives summary statistics for all tests except Written English. Table 9 presents correlations among tests and with sex, intelligence quotient, school attendance, and social status. Table 10 presents the same information for the test in Written English. Table 11 gives correlations for total and sub-scores in mathematics and problem solving. Comments include speculations about possible causes for the results, and suggestions for policy based on the results. (CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED164622

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

REPORT ON THE INTERMEDIATE EVALUATION PROJECT

I. Description and purposes of project
II. Comments on tests used in 1977
III. General findings and relationships
IV. Future Developments
V. Personnel involved with the project

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Mark Holmes

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

I. Description and Purposes of Project

The Intermediate evaluation project is a program developed by the Office of Field Development of OISE. It has been developed as a response to educators' requests that there be better information on just what children are learning in school.

Specifically, the purpose of the program was to develop measures of educational achievement that would (i) provide information to students, parents and teachers concerning the achievement of the individual;(ii) provide general information on the achievement of students overall, within a particular system, and within a particular school; and (iii) help school systems change their programs and methods in the light of specific information about strengths and weaknesses of their current curricula.

It was decided to measure achievement, broadly defined, at the eighth grade level as the vast majority of children have received common programs until that level. Young people typically make important decisions about their future education around the age of thirteen or fourteen. Most school boards felt that was the area of greatest need.

Tests in the following areas were developed for use in May 1977:

Mathematics (in French and English)
Problem solving (in French and English)
Science (in English only, there being no demand in French)
Attitudes (specifically: attitudes to school, reading, aesthetics, the environment and to service of others--in French and English)
French comprehension (French as a first language)
Written English (English as a first language)

Participation in the testing is shown in Table 1.

TM008 337

INTERMEDIATE TEST STUDENT PARTICIPATION BY BOARD Table 1

Board	Math Fr	Math Eng	Prob Solv Fr	Prob Solv Eng	Science	Attitudes Fr	Attitudes Eng	Francais	Written En
Brant County		392		341	506		249		191
Brant County Separate									363
Carleton Separate	124		126			124		121	
Central Algoma							99		387
Dryden District		377					384		
Dryden Dist Separate		53					51		
Elgin County		539							552
Essex County		91		28	143				623
Frontenac-Len Add Sep									245
Grey County		134		134					306
Huron County				186			1094		407
Kent County		478		381					1253
Lake Superior		129		126					
Lambton County Separate		208		130	136				
Norfolk County		81			78				
North Shore									149
Ottawa			101	467				143	441
Oxford County									411
Perth County		207		100	115		95		113
Sault Ste Marie		786		908	637		619		
Sault Ste Marie Separate									350
Stormont, Dundas Separate									96
Sudbury		769							378
Waterloo County Separate		24		27	101		55		87
Windsor							365		130
Total	124	4268	227	2828	1716	124	3011	264	6482

A decision had to be made as to whether the tests would be criterion referenced or norm referenced. There are advantages in both types of test. The decision was made that the tests would be essentially criterion referenced. Students would be measured according to their achievement of criteria of education as laid down by the Ontario Ministry of Education in its curriculum documents. The objectives were elicited from the documents, made more explicit, and circulated among teachers for validation. Questions were developed based on the objectives.

For each test three levels of performance were determined. A level of competence was designated. It was calculated from the questions on the assumption that an eighth grade student would be basically "literate" in that test area if that particular level of attainment was achievement. No assumption was made that a given percentage of students would or should reach the level (it is not a norm or mean). It is assumed that the vast majority of children could reach the level of competence if they were taught by an average teacher working very specifically towards the objectives on which the test was based. In fact, problem solving, as one example, is probably not taught very systematically and therefore it was not expected that a very high proportion of students would reach the level of competence. The second level, honors, was calculated to indicate the students had reached a good standing in the subject area of the test, one that should form a good basis for further study in secondary school. The third level, high honors, indicated that students had achieved a degree of mastery in the subject concerned. In Attitudes, three analogous levels were named--Acceptable Attitudes, Favorable Attitudes, and Highly Favorable Attitudes.

II. Comments on Tests Used in 1977

Mathematics

Table 2

Overall Results--Grade 8 (Anglophone) N = 2635

Sub-test	1 Computation	5 Measurement	3 Problems	2 Understanding	4 Algebra/Geometry	Total
Level of Competence	52	12	16	20	10	110
% of pupils this level						39
Honors	72	16	22	26	14	150
% of pupils this level						24
High Honors	80	28	28	44	20	200
% pupils this level						3
Provincial Mean Grade 8	54	12	17	27	15	125
(Provincial Mean Grade 7	46	11	14	24	12	107)

35% of the pupils taking the test failed to reach the level o competence. One obvious explanation is that the test was "too hard." In one se se, the test obviously was too hard--just because a significant proportion of st dents did not do very well. Our judgment is that the test was not too hard and that if the objectives of the course were given close attention, the vast majority of stud nts could reach the level of competence.

It will be noted that measurement, computation, and problems ere the worst answered sub-tests. Computation and measurement are very likely the areas most susceptible to direct instruction. Nevertheless, only 45% of the p pils could successfully answer a question like: 20 575 ÷ 149. The successful propo tion dropped to 38% when a division of a decimal number was required. Only 61% wer able to work out how much change was needed to pay a family's bus fare on an exact f re basis. Only 46% were able to calculate very simple percentages such as 20% of 6 0.

It is a common criticism of the "new mathematics" introduced n the 60's that computation is not systematically taught. One point of view is tha computation, in the age of pocket calculators, is not very important. That is a matter for school boards to decide. However, it would be less than honest to suggest that grade eight students compute well in the light of the above evidence. Whether r not it matters is a matter of judgment. At least, it would seem difficult to disp te the need to be able to work out how much change one needs to pay a bus fare.

Scores in measurement undoubtedly suffered as a result of ign rance of the metric system and instruction in that system should improve performance. easurement questions involving metric symbols were generally answered correctly by fewer than 30%, and fewer still if they involved calculation of area or solution of time distance and speed problems. Most mathematics texts at the intermediate level put lit le stress on systematic approaches to problem solving and this would generally ppear to be a neglected area. Problems involving percentages were particularly b dly done.

The other two sub-tests were much more adequately answered. 7 % understood the term common factor, 67% understood the idea of sets, 72% understood mathematical symbols. 38% could solve a simple equation and 48% could make numer ical substitutions.

The test's reliability is .8. The sub-tests, particularly the shorter ones, are considerably less reliable and should be used primarily as gener al indicators of group strengths and weaknesses. With the exception of motion ge metry, not included in the test, the test fits the new Ministry objectives rea onably well.

In particular, the new objectives place a greater emphasis on everyday mathematics(in the jargon "low mathematics") precisely the area in which students were least successful.

If the tests' objectives are the same as those of school boards, then the areas requiring greatest attention are: metric measurement--particularly problems (area, volume, speed); everyday problems; computation.

Problem Solving

Table 3

Overall Results--Grade 8 (Anglophone) N = 1594

Sub-test	1 Logic	2 Decision-Making	3 Correlational	4 Experimental	Total
Level of Competence	42	16	32	20	110
% of pupils this level					35
Honors	58	24	40	28	150
% of pupils this level					15
High honors	70	32	60	38	200
% of pupils this level					.3
Provincial Mean Grade 8	37	16	36	22	111
(Provincial Mean Grade 7	32	14	31	21	98)

Approximately half of the pupils taking the test failed to reach the level of competence. The determination of the level of competence in problem solving is much more difficult than in the case of mathematics. We do not have much experience of what average children can learn from average teaching. Indeed if problem solving is systematically taught one may well find that this test is too hard, or, for that matter, too easy. Suffice it to say that students reaching an honors level in this test should be able to handle problems they meet in school with reasonable facility. Students appear to do slightly better in correlational and experimental questions than in decision making and logic--perhaps because the first two are more often taught, in geography and science.

Logic appears to be particularly poorly understood, which is unfortunate given recent curricular stress on understanding concepts as distinct from acquiring knowledge. It might be worthwhile considering whether some teaching in the social sciences is

not promoting the development of opinions, as distinct from knowledge or understanding. Floyd Robinson, principal designer of the test, believes there is reasonable evidence that problem solving can be taught. The question then becomes one of priorities--how high a priority should be given to the teaching of problem solving?

Judgment of overall levels is difficult in problem solving until it is sytematically and more generally taught.

Science

Table 4

Overall Results--Grade 8 (Anglophone) N = 1044

Sub-test	1 Biology	3 Chemistry	2 Physics	4 Concepts	5 Man	Total
Level of Competence	48	7	20	18	17	110
% of pupils this level						41
Honors	58	10	32	28	22	150
% of pupils this level						36
High honors	72	15	51	34	28	200
% of pupils this level						1.6
Provincial Mean Grade 8	60	6	31	21	19	136
(Provincial Mean Grade 7	55	5	27	18	16	121 N = 667)

Whereas 35% of the pupils failed to reach the level of competence in mathematics, only 21% failed to reach the equivalent level in science. Was the test then "too easy"? We think not. However, the test is quite strongly biased towards biology--towards environmental studies and had the test been more biased towards pure science, particularly chemistry, the results would have been considerably worse.

The results are somewhat surprising, and contrast with those in mathematics. The "new mathematics" of the 60's brought about an emphasis on the meaning of mathematics, and have been criticized for leading to insufficient emphasis on computation and the mathematics of everyday life. In fact, pupils did particularly badly on computation and the mathematics of everyday life. The same new wave of the 60's brought an emphasis on science on the process and structure of science, with a de-emphasis perhaps on the traditional nature study. The science test results do not show any great strengths in scientific concepts, and young people's general knowledge

of scientific facts, particularly biological, is adequate.

General understanding of living things, plants and animals is high (over 80% getting such questions correct). 24% correctly answered a very tricky question on vertebrates and 36% knew how plants make food. Knowledge of the functions of the parts of the mammal is similarly high and 85% know the life cycle of the butterfly.

Basic knowledge in physics was also adequate although not as extensive as knowledge in biology. 39% know how thermometers operate and 34% have an elementary understanding of magnetism. There were no striking gaps in knowledge of this science.

Knowledge of chemistry was much more rudimentary. Less than one-third know the definition of elements, compounds, etc. Chemistry, rightly or wrongly, does not appear to be generally taught in the senior public school.

Understanding of concepts and the relationship of science to mankind, which one might have expected to be emphasized in science programs, is adequate but not remarkable. Only 25% understand the scientific concept of time. Only 59% could see that one cannot directly measure the circumference of a circle with a ruler (perhaps, though, they don't know what a circumference is). 40% could read charts and graphs and 23% could answer a quite difficult question on the nature of a scientific hypothesis. About half the pupils knew how to record the steps of a scientific experiment.

In general, students seem to have a fairly good basic knowledge of science, particularly biology and environment. Their understanding of scientific processes and of scientific method is much more scanty. The test seems to relate very closely to the new intermediate guidelines.

Attitudes

Table 5

Overall results--Grade 8 (Anglophone) N = 2223

Sub-test	2 School	1 Reading	4 Aesthetics	3 Environment	5 Others	Total
Acceptable attitudes	36	20	24	20	25	125
% of pupils this level						38
Favorable attitudes	44	32	36	28	35	175
% of pupils this level						41
Highly favorable attitudes	53	40	42	35	40	210
% of pupils this level						13
Provincial Mean Grade 8	47	28	36	30	31	173
(Provincial Mean Grade 7	46	28	37	29	30	169 N = 755)

Only 9% of the pupils had something less than acceptable attitudes in these five important dimensions. This is a reassuring discovery, although not a surprising one to those who have looked at research into pupils' attitudes to schools. It is a popular fiction perpetuated by the media that most children hate school. It is also interesting to note that the scores of Grade 7 pupils are virtually identical to those of Grade 8 pupils--the popular myth is that after enjoying elementary school children gradually get turned off in the senior elementary grades.

It may be objected that these statistics are invalid as children tend to give the answers that are expected. Certainly we cannot be sure that the answers to our questions are truthful. The press reports ignored the section of our June press release relating to favorable attitudes to school, although one report did highlight the bad news from the attitude test, that 29% wouldn't mind if Quebec separated from Canada. One reporter said in conversation that one could not expect children supervised by teachers to say they do not like school.

These are matters of judgment. There can be no doubt that some children gave answers that they thought would be acceptable. In the same way, there is no doubt that some children could have worked out some of the math questions but did not bother to. But is the proportion of dishonest pupils high enough to distort the results significantly? There can be no definite answer but it does seem improbable.

In the first place, children who dislike school or dislike reading tell their parents, their friends, and their teachers, so why would they not tell a questionnaire? In the second place, if pupils are generally giving the answers they think are expected, why do they respond so differently to different questions? For example, 89% think that education has some value for them but only 74% claim to enjoy school and 79% think their own school is a good one. There seems to be some discrimination at work. 77% enjoy leisure reading but only 48% wish they had more time for reading. It certainly does not appear that children are systematically giving the answers they think the testers want.

Only 32% think there are too many rules and regulations at school--were pupils really too terrified to say what they really think? This is consistent with anonymous questionnaires that have been administered in other parts of Canada--which found that a large minority of pupils actually think schools are not strict enough.

Attitudes to the environment are equally favorable--only 8% agreed that refundable bottle deposits are bad because of the trouble they cause. 18% think it would be

fun to chase wild animals on a snowmobile--a distressing finding perhaps, but it sounds honest.

In general, responses to questions on music, art and drama were favorable--schools don't seem to have made children dislike these activities. Similarly, most pupils have favorable attitudes towards their fellow men and to their country. Some teachers objected that the question asking if pupils had worked for others at school outside school hours was unfair for rural students who have to catch school buses. The complaint is as enlightening as the response. 50% of the pupils said they had not worked in such a manner. That would appear to be rather unfortunate. Helping to serve one's community is surely an important activity at school and it is a pity if bused pupils don't get the chance. Surely it is possible to give children useful opportunities to help others at noon hour and even in the minutes between arrival and start of school (some buses arrive as much as thirty minutes before school starts). Schools appear to have a stereotyped view of extra-curricular activities as consisting of formal sports starting after school finishes in the afternoon.

The statement that pupils have favorable attitudes to their country should be modified. Certainly there is no evident shame in being Canadian but the positive feeling does not run very deep. 29% think that the Maritimes and Quebec are not as Canadian as Ontario--would 29% of New Yorkers think that Texas is not as American as New York? 23% would not be very sorry to move to another country--is that the spirit of adventure, or just not a very deep involvement with Canada? 15% don't even think Canadians are pretty lucky compared with most other people, and 31% don't think there is much difference between Canada and the USA. If Canadians don't realize how fortunate we are, and if we don't think there is much difference between us and the USA, then there is not much hope for us to stay together as an independent country.

In general, the attitudes results are most favorable to the schools. The vast majority of students seem to enjoy school and to have developed healthy attitudes. The two exceptions seem to be service to others and a strong sense of attachment to their country.

French Comprehension

Table 6

Overall results--French Comprehension (Francophone)

Sub-test	1 Central Idea	3 Comprehension	5 Directions	2 Decoding	4 Use of Dictionaries	Total	N
Level of Competence (Gr. 8)	15	34	6	20	35	110	
% Gr. 7 reaching this level						39	62
% Gr. 8 reaching this level						41	58
% Gr. 9 reaching this level						47	142
Honors (Gr. 8)	21	43	8	30	48	150	
% Gr. 7 reaching this level						19	62
% Gr. 8 reaching this level						19	58
% Gr. 9 reaching this level						23	142
High Honors (Gr. 8)	29	63	16	38	54	200	
% Gr. 7 reaching this level						0	62
% Gr. 8 reaching this level						0	58
% Gr. 9 reaching this level						1	142
Mean (Gr. 7)	14	36	8	25	31	114	62
Mean (Gr. 8)	16	35	7	25	38	120	58
Mean (Gr. 9)	16	38	7	28	36	124	142

Results of this test can best be assessed in terms of individual classes in schools as the numbers involved are too small to make meaningful comparisons.

Obviously, with 45% of the Grade 9's, 40% of the Grade 8's, and 42% of the Grade 7's failing to reach the level of competence the pupils found the test difficult. The testing of Francophone students in an Anglophone environment poses severe difficulties. Should the level of competence be that which one would consider reasonable in a French speaking environment, related to the demand for studying and working in French? We felt that that standard would have been unreasonably high. After all, our other standards have been developed taking into account what could be achieved by an average teacher with average students working directly towards the objectives. But, how does one determine what a reasonable level is for Francophones in Ontario? And is what is reasonable in Ottawa reasonable in Welland? Certainly, we did not want to norm the test--in effect defining as reasonable what children actually do. Resulting levels were a compromise, below what we would consider reasonable for Quebecois, but above what is actually achieved in Ontario.

As the numbers of students are small, it would be difficult to make too many conclusions. One notes the lack of difference between the grades. As the students are in different schools and different boards one is hesitant to ask the obvious question--are the teachers aware of the levels of achievement of their students and do they teach systematically with the student difficulties in mind? Is there in fact considerable improvement over the grades, or does the Anglophone environment interfere very greatly?

Consideration might be given to a very systematic upgrading of vocabulary and expression using some second language techniques. For example, only 78% understand the word "poulain" although its meaning could be determined from the context. 57% understand the word "étui"--once again the context provided strong clues.

Table 7

Overall results--Written English

	Grade 7	Grade 8
Level of competence (Gr. 8)	-	110
Honors (Gr. 8)	-	150
High Honors (Gr. 8)	-	200
% reaching level of competence	55	37
% Honors	37	57
% High Honors	1	3
Mean	143	155
N	1890	4293
S.D.	23.4	24.1

Scoring of the written English was quite different from that of the other tests. Whereas, in the other tests, individual items were weighted so that a given, carefully selected group of correctly answered items would provide a score 110 (level of competence), a second group 150 and a third group 200, that method was quite impracticable in written English. The raw data in written English consisted of the sum of four scores on teacher marked essays, and two scores from scorers on each of two essays written in examination conditions in late April. Initially the scorers were calibrated to remove any systematic bias from consistently high or low scorers. Then the judgment was made by study of sample essays so that cut-off levels could be made denoting the level of competence and the level of high honors. The remaining essays were then distributed around those two fixed points.

As each group's (in most cases a class was a group) school scores from the teacher were standardized on the basis of the same group's performance on the examination, no comparison can be made between the school and the exam marks as a whole. Thus there is no way of saying whether the essays done in school were better or worse than those done in examination conditions at the end of the year. One can of course compare school and exam marks of an individual and see whether that individual did comparatively better on the teacher scored essays or on the final examination.

The level of competence in written English was defined in a way similar to that in the other tests. A student who scores 110 is minimally literate in written English

in the same way as the student who scores 110 in mathematics is minimally literate in number. In written English, 110 means that the student can communicate basic ideas in written English. A student who scores 200 is extremely literate--able to communicate quite complex ideas with clarity and grace. It is gratifying that only 3% of the Grade 8 students were not able to reach the level of competence, but disappointing that only the same percentage were able to reach the level of high honors. Generally the writing proved reasonably workmanlike, but lacking in interest, verve, variety, humor and life. (By the way, it deserves emphasizing that the fact that 3% received marks below 110 and 3% received marks above 200 is not a statistical artifact.)

One problem faced us in the written English that was not encountered in the other tests. The multiple scorers of the essays were instructed to mark the writing using numerals from 1 (very poor) to 10 (excellent)--thus assuming an interval scale and making a normal curve very likely. Our feeling is that the essays perhaps should not fit a normal curve. There are a great many mediocre but passable essays. That possibility is not reflected in the published scores.

Thus the level of competence is probably reasonably accurate (perhaps very slightly generous) but it may well be that the honors level should have been raised to 160 or 170. Whereas the other tests had three points fixed by close relationship with weighting for specific questions, the written English has only two fixed points.

One interesting point to notice is that, on the average, students improve by one-half a standard deviation between Grade 7 and Grade 8, a gratifying reflection on language teaching in eighth grade. Whereas a reasonable argument can be made that students' reading scores would increase between Grades 7 and 8 even without formal education, the same argument holds little force in written English. Most children do very little writing outside the school environment.

Peter Evans has written a much more complete report on the written English project and has described the kinds of error to which pupils are typically prone. Certainly there are structural, linguistic, and syntactical problems--but perhaps the greatest need is to encourage pupils to write with more feeling, more expression, more precision. The written English tests relate very well to the emphasis on written expression in the new Intermediate Guidelines.

III. Gene al Findings and Relationships

Table 8

Summary Statistics (for all students taking tests)

Test	N	Mean	Standard Deviation	Reliability	Standard Error
Mathematics	4268	118.5	39.2	.80	17.3
Problem Solvi ng	2828	105.5	34.0	.70	18.4
Science	1716	130.5	35.7	.84	14.1
Attitudes	3011	171.6	36.1	.84	14.5
No.1 Reading		28.0	12.7	.69	6.7
No.2 School		46.9	11.7	.68	6.3
No.3 Environm ent		29.4	6.8	.44	4.6
No.4 Aestheti cs		36.3	9.7	.64	5.6
No.5 Others		31.0	9.4	.41	6.8
French Compre hension	264	120.2	37.0	.71	19.8

The mos t important purposes of the project were to provide information on student achie vement to students and teachers. It was also hoped, as a secondary purpose, to i mprove understanding of the factors explaining student achievement. Particularly, we were interested to determine the relative effects of intelligence, social status , school attendance, teacher coverage of objectives, and school academic atmosphere, o n the achievement of Ontario pupils. Most of the available research is American and one cannot be sure about its complete applicability to the Canadian situation. I t was further hoped that schools and boards could be provided background data that wou ld help explain their achievement. That last question is dealt with separately ir specific, short reports to each board.

The fir dings must be treated with some caution as not all schools provided background ir formation (particularly I.Q.), and of those that did, not all provided information f or all students. We did not use student data in the correlation matrices if a school p rovided less than 50%--we assumed the remaining students would probably not constitut e a random sample. In calculating correlation coeffients, there are two ways of deali ng with missing data--list and pair deletion. The former method is preferable--o ne omits an entire list if some of the data are missing. With pair

deletion one only omits the information on single individuals for whom one item is missing. We adopted the latter procedure as the amount of missing data would have made the former method impracticable.

Table 9

Correlation Matrix--Major Variables

	Math	English	Problem Solving	Attitudes	Science	Sex	I.Q.	Attendance	Social Status
Math		.44*	.63*	.22*	.56*	-.03	.68*	.11*	.14*
English				.44*	-.02	.25*	.59*	.14*	.19*
Problem Solving				.34*	.59*	.03	.54*	.07*	.21*
Attitudes					.37*	.17*	.33*	.10*	.13*
Science						-.17*	.63*	.01	.23*
Sex							.05*	-.03*	-.01
I.Q.								.14*	.24*
Attendance									.08
Social Status									

* Significant at .01 level.

Table 5 shows few surprises. I.Q. correlates well with all four cognitive tests. It is interesting that it correlates no more highly with problem solving than it does with the other tests. We are interested to know whether the problem solving test measures something other than reading skill and intelligence. It is to be hoped that in next year's testing some schools will volunteer to administer reading, problem solving and will also provide I.Q. scores.

Social status explains much less variance in achievement than one usually expects. Possible explanations are: social status differences in Ontario do not influence achievement as much as they do in some other countries; our measures of social status (provided by schools) are not very accurate ones; social status differences in our population of schools (biased towards rural Ontario) are not as marked as they are in urban and suburban environments. In general, for our population, social status does not appear to be an important reason why some schools achieve better than others.

Attendance proved to be a rather poor predictor of achievement, with virtually no relationship in science and problem solving. Girls did somewhat better than boys in writing, and boys a little better in science, with no difference between the sexes in mathematics and problem solving.

We ran regression analyses of the significant independent variables (attendance, I.Q., and social status) and found that each factor remained fairly stable when the other factors were held constant. I.Q. explains 25% to 30% of the test variance (with the other variables held constant), social status 3% to 5%, and attendance about 1%.

Generally speaking, the degree to which teachers claim they emphasized various objectives (coverage) did not relate well with achievement. The correlation between coverage in math and achievement in math was .05, between coverage and problem solving achievement -.08, and between coverage in science and achievement -.1. There would appear to be two reasons for this finding. Firstly, teachers generally claim to cover nearly all our objectives--which is not surprising given that they were based on Ministry documents. Therefore, there was comparatively little variation in teachers' responses. Secondly, previous research finding significant relationships was based on teacher observations of test items, not teacher consideration of objectives. Where teachers see actual test items they know very well whether or not they have taught it (and perhaps where the students have learned it) and there may well also be an element of prediction in their responses.

The relationship between attitudes and achievement should not be misinterpreted. The fact that there is a tendency for children who like school, like reading etc., to do a little better than children who don't, does not mean that schools will improve their achievement if they make themselves more popular. Research generally shows that there is a slight relationship at the individual level between attitudes and achievement but that there is no relationship at the school level. Thus we were not at all surprised to find that some boards that came well above the provincial average in achievement also came below the provincial average in attitudes.

Written English

Table 10

Written English (Grade 8)--Correlation Matrix

	School Mark(Scaled	Test Mark	Final Mark	I.Q.	Social Status	Attendance
School Mark(Scaled)		.63	.94	.57	18	.14
Test Mark			.86	.48	15	.10
Final Mark				.59	19	.14
I.Q.					24	.14
Social Status						.08
Attendance						

Table 10 shows the scaled mark (scaled on the basis of the group performance on the test) correlates very slightly more highly with intelligence, social status and attendance than does the test mark--suggesting either that the classroom teacher is a very slightly better marker of written English than are the multiple scorers or that the classroom teacher is very slightly biased towards intelligent children from good homes. Either interpretation is reasonably flattering to the teacher. The teacher marks fairly well(once systematic bias is removed)--and if there is social bias it is extremely slight. There is no evidence at all either that there is a significant degree of social class bias on the part of teachers or that their marking is indiscriminate. The higher correlation between the school mark and the final mark than the test mark with the final mark is a statistical artifact--we weighted the school mark more heavily.

Mathematics and Problem Solving

Table 11

Mathematics and Problem Solving (Grade 8) Correlation Matrix

	Math.	1 Comp.	2 Understanding	3 Prob.	4 Alg/Geo.	5 Meas.	Prob.Sol.	1 Log.	2 D.M.	3 Corr.	4 Exp.	I.Q.	S.S.
Mathematics		.87	.75	.69	.63	.51	.63	.56	.38	.48	.30	.68	.14
Computation			.50	.46	.45	.33	.53	.46	.31	.41	.26	.58	.10
Understanding				.45	.41	.24	.51	.46	.29	.37	.25	.53	.09
Problems					.35	.28	.50	.45	.31	.38	.21	.52	.15
Algebra/Geometry						.22	.38	.37	.20	.25	.22	.40	.08
Measurement							.31	.25	.21	.28	.10	.30	.08
Problem Solving								.83	.59	.78	.56	.54	.21
Logic									.35	.45	.31	.50	.19
Decision Making										.31	.25	.34	.09
Correlational											.28	.37	.17
Experimental												.21	.11
I.Q.													.24
Social Status													

One caution to bear in mind when interpreting table 11 is that there is much more error in the sub-tests than in the total tests--and that alone accounts for the generally lower correlations. The generally high relationship between math scores and problem solving scores is interesting--it generally equals the relationships with I.Q. Problem solving sub-test 1 Logic is particularly noteworthy in this regard. It correlates with the math sub-tests as well as they do with each other although the problem solving items were virtually free of any mathematical calculation.(There were some questions requiring a little mathematics in problem solving but they were not in this sub-test.) When we checked the validity of our tests we looked at relationships between the students answering a particular item correctly and the proportion of the same students answering other items in the sub-test and total test correctly. If there was a strong relationship, particularly at the sub-test level, we decided the question was a good one. The individual items on the logic sub-test are interesting in that they generally correlated very well with the whole test, and only very slightly better with the sub-test. All this leads one to a tentative conclusion that logical skills may be very closely related to ability in mathematics and problem solving, although they are no more related with intelligence than other skills. Should schools be stressing the development of logic more?

Generally speaking, most schools only took one or two tests and even in those cases other information, such as I.Q., were sometimes lacking. That made it impossible for us to carry out the part of our research that would examine relationships at the school level as distinct from the individual level. It was at this level that we expected to find most relationships with respect to coverage and also the effects of academic atmosphere in the school. Academic atmosphere was calculated from items on the attitudes test and insufficient schools took the attitudes tests and the other tests to make valid comparisons.

IV. Future of the Project

The immediate future of the project is determined. The same tests will be offered in May 1978 in a modified or revised form. Exact equivalents of 1977 scores will be available for boards wishing to make comparisons.

In addition, several new tests are being developed and will be available in 1978. Present plans are for the development of new tests in: values, French as a second language, Canadian geography, Canadian history, Canadian Studies, health, physical fitness, reading and language. Tests will be available in French if there

is sufficient demand.

Considera tion is being given to the further use and development of the tests after 1978. They cou ld quite easily be normed (so that scores could be given in percentiles) and there has b een some demand for test availability in December or January instead of May. We wou ld be interested in receiving reactions to the test development so far and comments on its future.

V. Involveme nt in the Test

Those res ponsible for the Intermediate Evaluation Project are:

J. Clipsham (now with Niagara South Board) science
D. Ellis (OISE) mathematics
P. Evans (OISE Midnorthern) written English
M. Holmes (OISE) Project Director; attitudes
D. Jantzi (OISE) Project Coordinator
R. Lamerand (OISE) French comprehension
M-N Maillard (OISE) Coordinator French language schools and French language tests
F. Robinson (OISE Northeastern) problem solving
R. Stinson (OISE Western) science
R. Wolfe (OISE) statistical consultant.

1977 09 15

Mark Holmes

ADDENDUM

Note re var ables:

Social stat is was categorized on a scale from 1 (low) to 5 (high).

Attendance (excelle was categorized on a scale from 1 (poor attendance) to 4 nt attendance).

I.Q. was pr ovided, where available, from individual and group tests with a mean of 100.